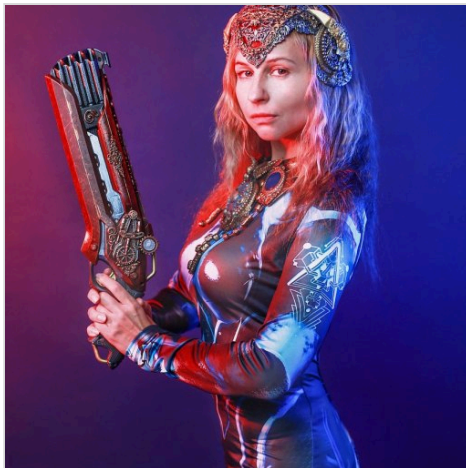


L U M B R E



Encoder-decoder y posicionamiento

Grado: 8

Materia: Transformers y Modelos de Atención · Inteligencia Artificial

11/10/2026

Antes de empezar

El transformer original combina un codificador que entiende la entrada y un decodificador que genera la salida, unidos por atención cruzada. Como la atención ignora el orden, se inyecta información de posición.

QUÉ APRENDERÁS

- Describir la arquitectura transformer codificador-decodificador y justificar la necesidad de la codificación posicional.

Encoder-decoder y posicionamiento

Devolver el orden a la atención

La atención es «ciega al orden»: permutar las palabras no cambiaría su salida. El posicionamiento y la estructura codificador-decodificador devuelven al transformer el sentido de la secuencia.

VIDEO

El decodificador del transformer

Vídeo en español que recorre capa a capa la arquitectura codificador-decodificador original.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=6JCVR2eCDe8>

Visión de conjunto

El codificador convierte la secuencia de entrada en representaciones contextuales ricas. El decodificador genera la salida un token cada vez, consultando al codificador.

Capas apiladas

- Tanto codificador como decodificador apilan varias capas idénticas.
- Cada capa refina la representación de la anterior.

- Profundidad significa más abstracción, pero más coste.

Bloque del codificador

Cada capa del codificador tiene autoatención (sin máscara, ve toda la entrada) seguida de una red de avance, con residuales y normalización en ambos submódulos.

Bloque del decodificador

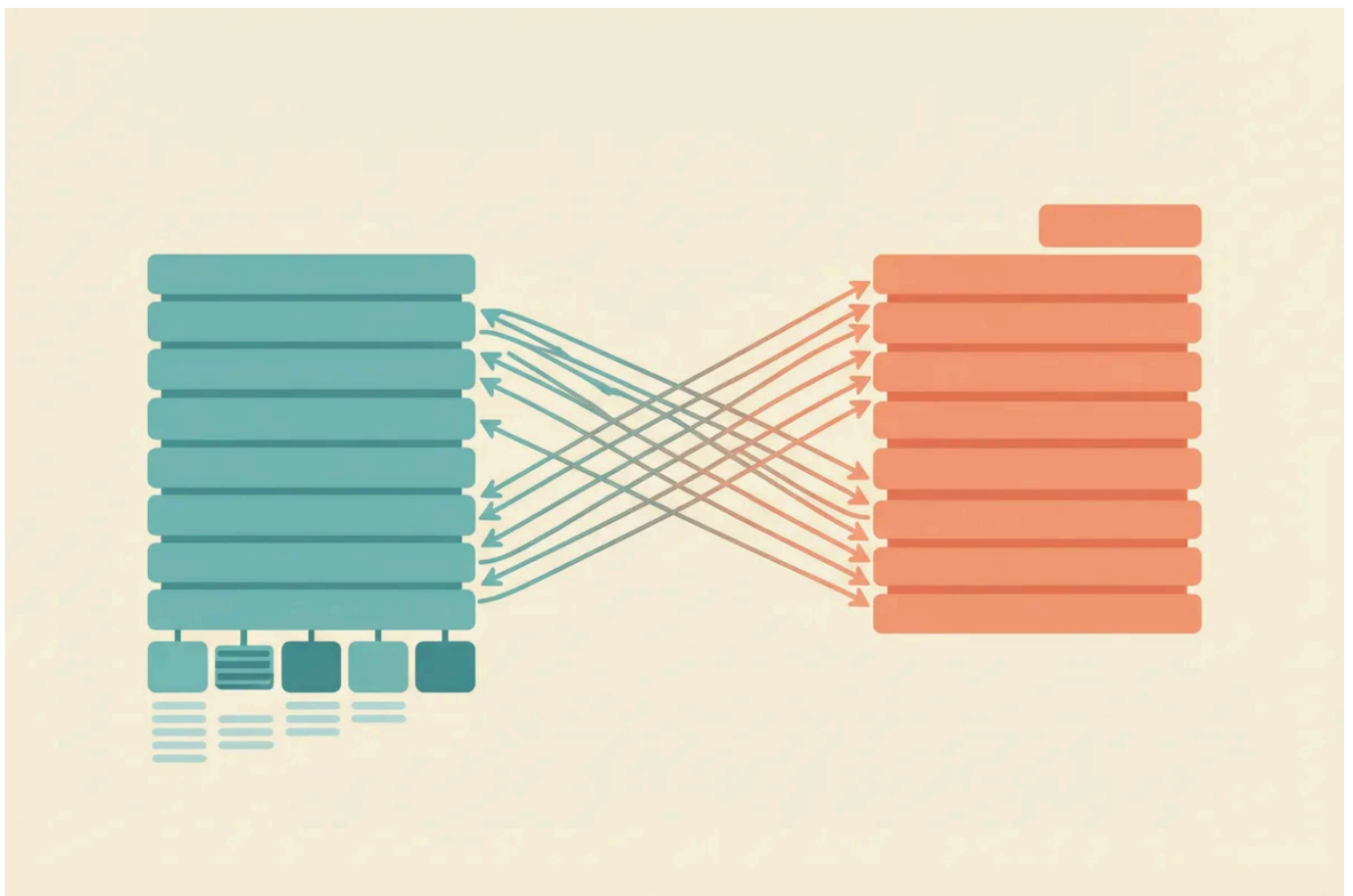
El decodificador añade una segunda atención, la cruzada: sus consultas miran a las claves y valores del codificador. Su autoatención va enmascarada para no ver el futuro.

Atención cruzada

Es el puente entre entrada y salida: en traducción, la posición que se genera consulta a toda la frase de origen mediante atención cruzada.

Red de avance

Tras la atención, un pequeño perceptrón multi-capa transforma cada posición por separado, aportando capacidad no lineal. Alterna con la atención capa a capa.



El codificador entiende la entrada, el decodificador genera la salida y la atención cruzada es el puente entre ambos.

Por qué hace falta posición

La atención trata la secuencia como un conjunto: no distingue «el perro muerde al hombre» de «el hombre muerde al perro». Sin señal de posición, el orden se pierde.

Posiciones senoidales

El artículo original inyectó funciones seno y coseno de distintas frecuencias, sumadas a la incrustación. Codifican posición absoluta y relativa de forma determinista.

Incrustaciones aprendidas

Alternativa común: aprender un vector de posición por índice, como una tabla. Suele funcionar mejor con datos abundantes, aunque no generaliza a longitudes no vistas.

Residual y normalización

Cada submódulo va envuelto en suma residual y normalización de capa, que estabilizan el entrenamiento de decenas de capas apiladas.

Entrada y salida en el decodificador

Durante el entrenamiento se usa docencia forzada: se alimenta la salida verdadera desplazada, paralelizando la generación de toda la secuencia.

Errores frecuentes

- Olvidar la codificación posicional y obtener un modelo que no distingue «el perro muerde al hombre» de su inversa.
- Esperar que las incrustaciones aprendidas generalicen a secuencias más largas que las vistas en entrenamiento.
- Confundir atención cruzada con autoatención: en la cruzada las consultas vienen del decodificador, no de la misma secuencia.

Ejemplo resuelto: traducir «the cat slept» al español

El codificador procesa la frase inglesa de una vez; el decodificador va generando la española token a token, mirándolo por atención cruzada.

1. El codificador autoatiende sin máscara las tres palabras de origen y produce sus representaciones.
2. El decodificador recibe el token de inicio y se autoatiende con máscara causal sobre su salida parcial.

3. Sus consultas se vuelcan por atención cruzada sobre todo el estado del codificador.
4. Emite «el» y lo anexa; el proceso se repite con «gato» y «durmió» como prefijos crecientes.
5. Si la posición «durmió» sale mal, revisas los pesos cruzados: delatan si el decodificador miraba donde debía.

¿Para qué sirve en la realidad?

Traductores automáticos, resumen de documentos largos y asistentes de voz con entrada y salida distintas son herederos directos de este diseño: un bloque entiende, el otro redacta y la atención cruzada los mantiene conectados.

1. ¿Por qué un transformer necesita codificación posicional explícita?

- ☐ porque el softmax no distingue signos
- ☐ porque la atención es invariante al permutar la secuencia
- ☐ porque los residuales borran el orden
- ☐ porque el decodificador no tiene atención

2. En el decodificador, la autoatención va enmascarada para impedir que cada posición vea tokens futuros.

- ☐ Verdadero
- ☐ Falso

ACTIVIDAD INTERACTIVA DISPONIBLE EN LUMBRE

3. Encoder-decoder



4. Ordena el paso de información en un transformer de traducción:

Elementos:

- el codificador procesa toda la frase de origen
- el decodificador se autoatiende enmascarado sobre su salida parcial
- las consultas del decodificador miran al codificador
- se emite el siguiente token de la frase destino

Codificar	. . .
Autoatender	. . .
Cruzar	. . .
Generar	. . .

Transformer [https://en.wikipedia.org/wiki/Transformer_\(deep_learning_architecture\)](https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture))

Arquitectura codificador-decodificador basada solo en atención.

5. Une cada pieza del transformer original con su función.

5.	procesa toda la entrada y la resume en representaciones contextuales	-----
5.	genera la salida un token cada vez con máscara causal	-----
5.	puente que deja al decodificador consultar al codificador	-----
5.	señal sumada que devuelve al orden a la atención	-----

Opciones: codificador · decodificador · atencion cruzada · codificacion posicional

Entender y generar por separado

Transformer codificador-decodificador

Codificador

autoatencion sin mascara

toda la entrada

Decodificador

autoatencion causal

atencion cruzada

Posicion

senoidales

aprendidas

Estable

residual

normalizacion

red de avance

El mapa mental interactivo es interactivo en la versión online.



Modelos secuencia a secuencia <https://en.wikipedia.org/wiki/Seq2seq> La familia de arquitecturas que transforman una secuencia en otra.

6. Sin codificación posicional, «a sigue a b» y «b sigue a a» darían la misma representación. ¿Qué dos estrategias de posicionamiento compararías y cuál generaliza a secuencias más largas?

Referencia curricular

Grado:	8
Materia:	Transformers y Modelos de Atención · Inteligencia Artificial

Versión interactiva completa (con actividades, videos y autoevaluación):



<https://univ.lumbre.lat/object/cdia-tranf-003-encoder-decoder-y-posicionamiento>

Material educativo generado desde Lumbre · lumbre.lat